17

16

15

14

13

12

11

10

9

8

7

6

5

4

3

2

1

[www.CrosswordWeaver.com](http://www.CrosswordWeaver.com/)

# ACROSS

**1** A tool used to handle missing values by filling them in a straightforward manner.

**6** General method of converting data from one format to another, often categorical to numerical.

**8** Measures how data points spread out from their average.

**9** When one category greatly outnumbers another in a classification problem, this issue arises.

**10** A method to reduce words to their root form, often used in natural language processing.

**12** The process of converting text into numerical format for machine learning tasks.

**13** When a model's predictions are consistently off in one direction, it has this issue.

**14** A visual representation often used to display the relationship between two variables.

**16** The process of creating new input attributes or modifying existing ones to improve model performance.

# DOWN

1. Popular Python library for a range of machine learning algorithms; abbreviated.
2. A single unit, often a word or number, resulting from the process of breaking down text data.
3. This technique turns categorical data with a specific order into numeric format.
4. A statistical measure indicating the degree to which two variables move in relation.
5. Python library favored for data wrangling and analysis.

**7** This technique transforms categorical data into a format where each category is represented by a unique binary column.

**11** A technique to increase the number of instances in the minority class by replicating them

**15** This value represents the average of a dataset.

**17** Initial exploratory step in data analysis; abbreviated

**Answers:**

**ACROSS**

1. **SIMPLEIMPUTER**
2. **ENCODING**
3. **VARIANCE**
4. **CLASSIMBALANCES**
5. **STEMMING**
6. **VECTORIZATION**
7. **BIAS**
8. **SCATTERPLOT**
9. **FEATURECONSTRUCTION**

**DOWN**

1. **SKLEARN**
2. **TOKEN**
3. **ORDINALENCODER**
4. **CORRELATION**
5. **PANDAS**
6. **CATEGORYENCODER (ONE-HOT ENCODING)**
7. **UPSAMPLING**
8. **MEAN**
9. **EDA**